**Testimony by Nathaniel Good and Aaron Krekelberg**
**University of California, Berkeley**
**School of Information Management Systems**
**And**
**University of Minnesota**
**Office of Information Technology**
**Before the**
**House Committee on Government Reform**

Good morning Mr. Chairman and Committee members. Thank you for the opportunity to appear before you today. My name is Nathaniel Good. I am a graduate student at the University of California, Berkeley in the School of Information Management and Systems. My colleague, Aaron Krekelberg, is the Chief Web Architect at the University of Minnesota's Office of Information Technology. It is an honor to be here today to present to the committee and discuss the results of a study we performed on usability and privacy of the KaZaA peer-to-peer file sharing network.

## Goals of the Study

The primary goal of our study was to demonstrate that good user interface design is an essential part of designing an application that is secure and preserves users' privacy. By exploring how private information could be exposed by miscommunication between the user and the application, we hoped to illustrate how important it is to develop and incorporate human-computer design principles into the process of creating applications that could potentially leave users' data exposed. We also hoped to draw attention to the larger, more general problem of creating safe user interfaces for all types of continuously connected, networked systems that store and share users personal and private information.

## Summary of Study Results

In this study, we determined through both user studies and analysis that the KaZaA applications interface had several critical flaws that may contribute to participants' misconfiguring the application and thus inadvertently sharing their private and personal information. In the user study we conducted, only 2 of the 12 participants were able to correctly determine that the installation they were given was sharing all files on their hard drive. We conducted a survey with 12 participants and asked them to identify the types of files that could be shared using a P2P network (such as word documents, financial information, spreadsheets, music files, etc.). From the survey, we discovered that 9 out of the 12 assumed incorrectly that only certain types of files could be shared, rather than all files and file types on their hard drive.

We also conducted a study to determine how many other users unique inboxes we could find from our single KaZaA installation. By using this approach, we hoped to examine how a person on KaZaA could possibly search for others private information on the network without having to have any sophisticated tools or knowledge. Using this approach we were able to find 150 unique users inboxes in 12 hours, and almost 1000 users inboxes in a week.

In addition, we ran a dummy client sharing files that were disguised as personal files such as “credit cards.xls” and the email file “inbox.dbx” to determine if other KaZaA users were searching for and downloading these files from other users. Over 24 hours, we discovered that four unique users had downloaded “credit cards.xls” and two unique users had downloaded “inbox.dbx”.

## Summary of Conclusion and Findings

It is our opinion that the problems we discovered with the KaZaA interface are not intrinsic to P2P in general, nor are they a reflection of an underlying security weakness in P2P systems that “causes” users to share files without their knowledge. We feel that the problems we describe in our report can be adequately addressed by educating users about P2P and networking in general, and more importantly, improving the user interface for the KaZaA application following the guidelines described in our report. The default settings should recognize that all files are not created equal, and some file types shouldn’t be available for sharing by default, such as email, excel spreadsheets, tax returns etc. To provide the maximum protection for users sharing files, the default settings should be configured to prevent sharing of potentially harmful files and file types. In addition, any modifications to these settings should be easily recognizable for others who may not have configured the application, but share the computer on which it is installed.

## Background of the KaZaA study

Several months prior to our initial study, we became increasingly aware of personal files such as email, spreadsheets and financial documents appearing in search results on KaZaA. We initially assumed that the results were limited to isolated cases, but after several months were convinced that the problem was larger than we initially suspected. An initial investigation of the user interface of KaZaA, along with anecdotal accounts from several KaZaA users, led us to believe that confusion around the user interface could account for users inadvertently sharing more information than they intended, including the personal and private information we were seeing on the network. We decided to run a study to test our hypothesis.

KaZaA was interesting from a research perspective because it is widely used, has user interface issues that could compromise users privacy, and has grown rapidly from a small knowledgeable user base to a large user base with many users of very different backgrounds and levels of computer experience. Unlike previous P2P file sharing services such as Napster, KaZaA allowed users to not only share music files in the popular mp3 format, but any other kind of file as well. Also, despite a relatively safe default installation, there were many people sharing personal information without their knowledge. This suggested that a significant number of people had been misconfiguring the application after the installation had occurred. For this reason, we saw this as a problem with the applications usability, and chose to use techniques from human computer interaction to analyze it.

## What is Usability and Human Computer Interaction?

Human Computer Interaction is an interdisciplinary field that merges fields such as computer science, cognitive science and design. Its primary goal is to reduce the friction between humans and machines, and create a means for people to use machines as intuitively as possible.

One can think of Human Computer Interaction in terms of a highway system. A highway is designed to take people where they need to go, quickly, safely and efficiently. If there are confusing road signs, people may miss exits and have trouble getting where they need to go. If there are ill-designed roads that require people to jump across many lanes to exit, or have sudden curves or blind corners, the effects can be more than just irritating; they can be deadly. One can imagine several approaches to fixing poorly designed roads. One can put up signs alerting drivers to the dangers or changes, and hope that they read them. This approach could be considered one of education. The other approach is to try to redesign the road altogether, which can be quite costly. Human Computer Interaction is a discipline dedicated to ensuring that users have "smooth" rides when working with applications, improving existing applications that may currently be "bumpy" or frustrating for users, and assisting in redesigning interfaces and interactions that could have serious negative consequences.

It is important to explain the difference between this view and views traditionally discussed on security and privacy. When security breaches are typically described in the common press, they are described as errors or vulnerabilities in the program's code which allow attackers to take advantage of these mistakes and compromise the system. Typically, these kinds of errors can be corrected or "patched" by writing new code that fixes the problem, and then having the users download and install the "patch", thus plugging the security hole. For problems that exist with the user interface, it is not as simple as writing a patch. Adding more security in the form of data encryption or other technical measures will not help with misconfiguration problems or address problems with miscommunication. Eventually, the data being protected by such measures has to be unencrypted and handled by a user, and it is at this point that the system must help guide the user into making the correct choices and help prevent them from "shooting themselves in the foot" and making fatal mistakes. To fix these kinds of issues, the software creators need to rethink, test and redesign the user interface to properly address the problems.

## Details of the KaZaA Study

For our study we decided to look at whether (to the extent that we could measure) sharing personal files was a problem on the KaZaA network, whether other users knew this and were taking advantage of this a problem, and whether confusion with the user interface and assumptions about file sharing could be a cause of this problem.

### *Can I find other users' private information?*

For this question, we wanted to search for unique users who were sharing files that were personal in nature. A very personal file is ones email file. People generally do not want

strangers to read their email, so if people were sharing this file then we could assume that they might also be sharing other files that were private. We chose to search for the file "inbox.dbx" because it is common on all Windows machines, which is currently the only operating system that KaZaA supports. It also was a good choice because it typically resides in a folder that contains other private files, which people would not want to share. We ran test queries, and for each test query used the KaZaA function to "search for more files from this user" to see the other files that the user was sharing to confirm that they were sharing more than just the inbox.dbx file. In 19/20 cases, this assumption was correct. In the one case it wasn't, the user was only sharing a suspicious collection of many inboxes.

**Results**

For our initial study, in a 12 hour period we were able to find 156 distinct email inboxes. In a later study performed this year, over a 7 day period we were able to find approximately 1000 distinct email inboxes. In the first study, we looked more closely at a subset of 20 users and found that in addition to exposing files other than "inbox.dbx", 9 users had exposed their web browser's cache and cookies, 5 had exposed word processing documents, 2 had exposed data from financial software and 1 user had files that belonged in the system folder for Microsoft Windows.

### *Are other users' downloading KaZaA users personal files?*

For this question, we were interested in determining if other users on KaZaA were aware of some users sharing private information, and were taking advantage of this by downloading these files. To test this, we setup a KaZaA client to share personal and private files such a spreadsheet called "credit cards", and the email file described earlier, "inbox.dbx". We let our "honeypot" run for 24 hours and looked at the files downloaded over that period of time.

**Results**

From our dummy server, we received a total of four downloads from four unique users for an Excel spreadsheets named "Credit Cards.xls" and four downloads from two unique users of an Inbox.dbx file for our initial study. The second follow up study we performed this year had similar results for both file types.

### *Is the interface confusing users and does it match their assumptions?*

For this question, we created a user study to test if users could determine what files were being shared on a KaZaA installation, and if the problems we found in the initial interface analysis contributed to this confusion. In addition, we wanted to learn about the assumptions our users had about the types of files that could be shared on P2P file sharing systems, and how much experience they had with P2P. We had 12 users run through our task and answer a short survey on their computer experience, P2P experience and assumptions on the types of files that could be shared on P2P networks.

**Results**

10 of the 12 users had used file-sharing programs, and all were considered "experienced" computer users by the standard QUIS metric of greater than 10 hours of computer time a

week. Of the 12 users, only 2 correctly identified that KaZaA installation had been set to share all files on the hard drive. In addition, only 2 users correctly indicated that all types of files could be shared over a P2P network. 9 of the 12 users believed that only multimedia files such as music, video and pictures could be shared.

## Limitations of the KaZaA study

It is important to note what we did not study. We did not do a study of what percentage of files on the KaZaA network were personal files. The KaZaA P2P network is encrypted, and although reverse engineering the protocol is feasible, our understanding is that it is not currently allowed under the existing DMCA regulations, and also in the KaZaA user agreement. In addition, even if we were allowed to reverse engineer the protocol, the distributed decentralized nature of the network would make it difficult to look at it in its entirety. However, if we were allowed to reverse engineer the protocol we would be capable of examine the network contents and traffic in greater detail.

Because of these imposed limitations on our ability to conduct a more thorough probe of the KaZaA network, we were limited to automating the KaZaA user interface to perform out searches. A disadvantage of this approach is that it prevented us from knowing how much of the network we are searching at any given time. In addition, KaZaA's distributed "super-node" architecture is such that there is no guarantee that computers will connect to the same part of the network at any given time. For example, two computers may be physically next to each other, but would see completely different search results because they would be connected to different supernodes.

In addition, we did not perform a full scientific study on why users were sharing personal information. We could not speculate on all of the various reasons users would want to change their default settings, although we knew from our data that they were indeed modifying the settings and were not aware of the implications. Our initial goal was to describe how this could happen, given the anecdotal evidence we had from KaZaA users and the types of files we saw being shared. By analyzing this information, we determined that the types of files being shared were similar to files that one would find in system folders, document folders, program folders and in some cases, indicative of users sharing an entire hard drives' contents. Conversations with KaZaA users who were sharing this information and who responded to our requests confirmed that they were sharing these without their knowledge. For this reason, we hypothesized that configuration issues could account for users inadvertently sharing personal files, and we chose to concentrate on the user interface issues.

We would also like to state that during the course of the study, we did not download any files from users. Although it may have been legal, we felt it was not ethical to take this information from users. The types of files being shared, as well as comments from others who did download these files convinced us that some users were indeed sharing their private and personal information.

## Conclusions

Since the publication of our first study, KaZaA has responded by providing an explanation of how to configure the program on their website, although they have yet to modify the user interface. We are hopeful that by providing the information in our report and offering suggestions for improvement, KaZaA will take measures in the near future to redesign the most serious users interface problems we discovered.

The problems we describe are very much part of a larger, more general problem that applies to all networked systems that store and share users' personal and private information. The problems we described in the report could also exist in email applications (as reported in a related paper on usability and security by Whitten and Tygar), knowledge sharing applications and other types of applications that have sensitive information managed by users on continuously connected networks. We see our work in the context of a new and emerging interest in the field of Human Computer Interaction on providing secure and usable user interfaces to help users manage the complexities of access control for private, semi-private and public information. As the world becomes more networked, and devices and means for sharing and gathering personal information proliferate, work in this area is central to the design of applications that support peoples' privacy and security in a networked world.

Thank you very much for allowing us to present here today.